How Good Is a Broyden–Fletcher–Goldfarb–Shanno-like Update Hessian Formula to Locate Transition Structures? Specific Reformulation of Broyden–Fletcher–Goldfarb–Shanno for Optimizing Saddle Points

JOSEP MARIA ANGLADA,[1] JOSEP MARIA BOFILL[2]

[1] *C.I.D–C.S.I.C., Jordi Girona Salgado 18-26, E-08034 Barcelona, Catalunya, Spain*
[2] *Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès 1-11, E-08028 Barcelona, Catalunya, Spain*

Received 28 March 1997; accepted 8 September 1997

ABSTRACT: Based on a study of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update Hessian formula to locate minima on a hypersurface potential energy, we present an updated Hessian formula to locate and optimize saddle points of any order that in some sense preserves the initial structure of the BFGS update formula. The performance and efficiency of this new formula is compared with the previous updated Hessian formulae such as the Powell and MSP ones. We conclude that the proposed update is quite competitive but no more efficient than the normal updates normally used in any optimization of saddle points.
 J Comput Chem 19: 349–362, 1998

Keywords: Broyden–Fletcher–Goldfarb–Shanno formula; update Hessian formula; transition structures; optimization; saddle points

Correspondence to: Dr. J. M. Bofill; e-mail: jmbofill@canigo.qo.ub.es
Contract/grant sponsor: Spanish DGICYT; contract/grant number: PB95-0278-C02-01

 CCC 0192-8651 / 98 / 030349-14

Introduction

The study of potential energy hypersurfaces and specifically the location of their stationary points are crucial in modern theoretical chemistry. The most important stationary points are minima (i.e., equilibrium structures) and first-order saddle points (i.e., transition structures or TS). In general, the algorithms for locating this class of stationary points are based in the quasi-Newton–Raphson (QNR) type technique.[1] Basically, the methodology consists of defining at each step of the optimization process a quadratic model for the expansion of the energy function around some point of the hypersurface. The quadratic model is built by computing only the energy and the gradient at this point. The Hessian matrix or its inverse is approximated by updating it at each iteration rather than evaluating it exactly. This represents a big savings of computing time, because in general the second derivatives of the energy are much more expensive to compute than the gradient. Due to its general efficiency in achieving convergence and the relative low cost of the computation, the QNR methodology is a powerful tool to locate and optimize stationary points.

The most important point in any QNR algorithm is to find a good formula for updating the Hessian matrix or its inverse at each iteration. This updating formula technique was introduced by Davidon[2] and modified by Fletcher and Powell[3] and is often called the Davidon–Fletcher–Powell (DFP) method. Variants of this updating formula were developed by Broyden,[4] Fletcher,[5] Goldfarb,[6] and Shanno[7]; this most popular BFGS formula is the standard Hessian updating formula used in the QNR algorithms to minimize molecular geometries.

The mathematical nature of these update formulae can be analyzed from different points of view. In a short but beautiful article, Huang[8] derived and unified a family of update formulae, even for a nonsymmetric Hessian. His derivation was based on the preconditioned conjugate directions algorithm[9] applied to quadratic functions. This family of update Hessians (or its inverse) is called the Huang family. In principle, any member of the Huang family can be applied to locate maximum and saddle points of any order for quadratic and nonquadratic functions. However, for optimizing saddle points of nonquadratic functions, the Huang family of formulae present a very low convergence, in general. Greenstadt[10] showed that the update formulae can be derived from a minimization of weighted errors where the weights are positive definite. Based on these conclusions, Bofill and Comajuan[11] observed that when many update Hessian formulae members of the Huang family are applied to locate and optimize a saddle point, this leads to negative weights. However, this is not the sole drawback of the Huang family because they generally modify the approximate transition vector too much that is associated with the desired TS from iteration to iteration, producing a deterioration of the optimization process or a lack of convergence.[11]

To optimize a saddle point structure using a QNR algorithm, the most widely used update formulae are the Powell one[12] and its generalization, the so-called Murtagh–Sargent–Powell (MSP) family.[11,13,14] The use of the Powell formula for this type of optimization was first suggested by Simons et al.[15]

When applied to a quadratic function, the Powell and MSP update formulae[11–14] both present a limitation with respect to the BFGS update: they do not preserve the so-called conjugancy property[1,16] except when the conjugate search directions are mutually orthogonal. It is important that the update formula possesses the conjugancy property because it ensures that the QNR algorithm converges for a quadratic function in at most $n + 1$ searches, where n is the number of variables.[1,16] This result can be applied when the function to be optimized is not quadratic, because near a stationary point the function presents quadratic behavior. Consequently, the update formulae that preserve the conjugancy property without any limitation, like the BFGS, in general are more efficient in geometry optimizations of local minima than others like the Powell or MSP formulae, which are more restricted with this property. In locating and optimizing TS, one has the following dilemma: BFGS is the best update for optimizing minima but surprisingly it is not a good update for optimizing saddle points. The following questions arise. What is the reason for the low efficiency in the optimization of saddle points? Is it possible to reformulate the BFGS-like update formula to optimize saddle points in an efficient way? This article analyzes these questions.

Mathematical Background

QNR METHOD

The QNR method[1, 16] consists of first quadratically expanding around $\mathbf{x}^k$ a function $f(\mathbf{x})$ that depends on n real variables and is at least twice continuously differentiable:

$$f(\mathbf{x}^k + \Delta\mathbf{x}^k) \approx q^k(\Delta\mathbf{x}^k) = f(\mathbf{x}^k) + (\mathbf{g}^k)^T \Delta\mathbf{x}^k + \frac{1}{2}(\Delta\mathbf{x}^k)^T \mathbf{G}^k \Delta\mathbf{x}^k, \quad (1)$$

where $\Delta\mathbf{x}^k = \mathbf{x}^{k+1} - \mathbf{x}^k$, $\mathbf{g}^k$ is the gradient vector, $\mathbf{G}^k$ is the Hessian matrix, and $q^k(\Delta\mathbf{x}^k)$ is the quadratic approximation of $f(\mathbf{x}^k)$ for iteration k. The Hessian matrix $\mathbf{G}^k$ of eq. (1) is substituted by a symmetric nonsingular matrix $\mathbf{B}^k$, which is updated from iteration to iteration. Sometimes this update is made directly on the inverse of the $\mathbf{B}^k$ matrix, $\mathbf{H}^k$. In each iteration generally the correction $\Delta\mathbf{x}^k$ optimizes $q^k(\Delta\mathbf{x}^k)$, but sometimes $\Delta\mathbf{x}^k$ is chosen such that it optimizes $q^k(\Delta\mathbf{x}^k)$ in a selected neighborhood of $\mathbf{x}^k$.

In principle, if the $\mathbf{B}^k$ or $\mathbf{H}^k$ matrices have the desired eigenvalue spectra, that is, the number of positive and negative eigenvalues are correct, the QNR algorithm should converge to a stationary point such that its Hessian matrix $\mathbf{G}^*$ possesses the desired eigenvalue spectra at this point.

To locate the optimize saddle points, the use of a selected neighborhood during the optimization converts the QNR method to a very efficient one.[13, 17–19] This modification has been called a quadratic approximation,[18b] the trust region image method[18a] (QA/TRIM), or the restricted QNR[13] (RQNR). We note that other methods exist to compute the optimum step $\Delta\mathbf{x}^k$, like the so-called rational function optimization,[20] but they will not be discussed here.

We are interested in locating and optimizing TS. In this case, if one updates the approximated Hessian rather than its inverse, it is necessary to invert this matrix by diagonalization so the size and the sign of the eigenvalues can be controlled.[21] Due to this feature, we present only the update of the Hessian matrix, $\mathbf{B}^k$, rather than its inverse $\mathbf{H}^k$. In order to simplify the notation, hereafter the superindices will be dropped; that is, we take $\mathbf{B}$ for $\mathbf{B}^k$, $\mathbf{B}'$ for $\mathbf{B}^{k+1}$, and $\Delta\mathbf{x}$ for $\Delta\mathbf{x}^k$.

HUANG FORMULA'S TO UPDATE APPROXIMATE HESSIAN MATRICES

As pointed out in the Introduction section, Huang[8, 16] obtained a general formula for updating the Hessian matrices by comparison with the preconditioned conjugate direction method.[9] The formula is

$$\mathbf{B}' = \mathbf{B} + \rho\frac{C_1\Delta\mathbf{g}\Delta\mathbf{g}^T + C_2\Delta\mathbf{g}\Delta\mathbf{x}^T\mathbf{B}}{C_1\Delta\mathbf{g}^T\Delta\mathbf{x} + C_2\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}} - \frac{K_1\mathbf{B}\Delta\mathbf{x}\Delta\mathbf{g}^T + K_2\mathbf{B}\Delta\mathbf{x}\Delta\mathbf{x}^T\mathbf{B}}{K_1\Delta\mathbf{g}^T\Delta\mathbf{x} + K_2\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}}, \quad (2)$$

where $\Delta\mathbf{g} = \mathbf{g}^{k+1} - \mathbf{g}^k$ and ρ, C_1, C_2, K_1, and K_2 are parameters. Normally, ρ is taken as equal to 1. Selecting the other parameters conveniently one gets the well-known update formulae such as the Murtagh–Sargent (MS),[22] dual Greenstadt,[10] DFP,[2, 3] or BFGS.[4–7] Taking $\rho = 1$, $C_1 \neq 0$, $K_2 \neq 0$, and $C_2 = K_1 = 0$ we get the well-known BFGS update formula for the $\mathbf{B}$ matrix, hereafter labeled by $\mathbf{B}'_{\mathrm{BFGS}}$.

$$\mathbf{B}'_{\mathrm{BFGS}} = \mathbf{B} + \frac{\Delta\mathbf{g}\Delta\mathbf{g}^T}{\Delta\mathbf{g}^T\Delta\mathbf{x}} - \frac{\mathbf{B}\Delta\mathbf{x}\Delta\mathbf{x}^T\mathbf{B}}{\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}}. \quad (3)$$

Equation (2) can be seen also as an approximation of the Taylor series expansion of the Hessian $\mathbf{G}$ at the point $\mathbf{x}^k$, where the second and the third term of eq. (2) summarize the error due to the truncation. In other words, eq. (2) may be written in a compact form as

$$\mathbf{B}' = \mathbf{B} + \mathbf{E}, \quad (4)$$

and it can be seen as an approximation to the Taylor series expansion of $\mathbf{G}$ (see ref. 23),

$$\mathbf{G}' = \mathbf{G} + \mathbf{F}\Delta\mathbf{x} + \cdots, \quad (5)$$

where $\mathbf{G}'$ is the Hessian matrix at $\mathbf{x}^{k+1}$ and $\mathbf{F}$ is the tensor of the third derivatives at $\mathbf{x}^k$. To update the inverse Hessian matrix one uses an equation that is equivalent to eq. (4), that is, $\mathbf{H}' = \mathbf{H} + \mathbf{Q}$, where the $\mathbf{Q}$ matrix plays the role of the $\mathbf{E}$ matrix.[16]

Finally, we note that the BFGS formula can be reformulated with $\rho = -1$ to optimize saddle points starting with a positive definite $\mathbf{B}$ matrix. However, this has little practical value because this is only true in the strictly quadratic case.[8]

VARIATIONAL BASIS OF UPDATE HESSIAN MATRICES

Greenstadt[10] was first to apply a variational principle to obtain update formulae, but it was applied to the **H** matrix rather than the **B** matrix. Here we briefly summarize this variational principle, but we apply it only to the **B** matrix.[11] The QN condition is defined as[1,16]

$$\Delta \mathbf{g} = (\mathbf{B} + \mathbf{E})\Delta \mathbf{x} = \mathbf{B}'\Delta \mathbf{x}. \tag{6}$$

This condition is an approximation of the first-order Taylor series expansion of $\mathbf{g}^k$ with respect to the step vector $\Delta \mathbf{x}$ conveniently rearranged. Greenstadt[10] evaluated the **E** matrix by minimizing the following weighted Euclidean norm,

$$N_{\mathbf{W}}(\mathbf{E}) = \mathrm{Tr}(\mathbf{WEWE}^T), \tag{7}$$

subject to the QN condition, eq. (6), rearranged as

$$\mathbf{j} = \mathbf{E}\Delta \mathbf{x} = \Delta \mathbf{g} - \mathbf{B}\Delta \mathbf{x}, \tag{8}$$

and the symmetry condition $\mathbf{E} - \mathbf{E}^T = 0$. In eq. (7) Tr indicates the trace and **W** is a symmetric and positive definite matrix. The problem is formulated using the following Lagrangian function,

$$\Phi(\mathbf{E}, \mathbf{p}, \mathbf{L}, \mathbf{W}) = \frac{1}{2}\mathrm{Tr}(\mathbf{WEWE}^T) + \mathrm{Tr}[(\mathbf{E}\Delta \mathbf{x} - \mathbf{j})\mathbf{p}^T] + \mathrm{Tr}[\mathbf{L}(\mathbf{E} - \mathbf{E}^T)], \tag{9}$$

where the **p** vector and **L** matrix are the Lagrangian multipliers. With these considerations the formula to update the **B**′ matrix is

$$\mathbf{B}' = \mathbf{B} + \mathbf{E}^* = \mathbf{B} + \mathbf{j}\mathbf{u}^T + \mathbf{u}\mathbf{j}^T - (\mathbf{j}^T\Delta \mathbf{x})\mathbf{u}\mathbf{u}^T = \mathbf{B} + [\mathbf{j} \quad \mathbf{u}]\begin{bmatrix} 0 & 1 \\ 1 & -(\mathbf{j}^T\Delta \mathbf{x}) \end{bmatrix}[\mathbf{j} \quad \mathbf{u}]^T, \tag{10}$$

where the $\mathbf{E}^*$ matrix is the solution of eq. (9), $\mathbf{u} = (\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x})^{-1}\mathbf{M}\Delta \mathbf{x}$, and $\mathbf{M} = \mathbf{W}^{-1}$. In Appendix A we give a derivation of the most general inverse to eq. (10). It is important to emphasize that this variational principle does not say anything about the form of the **W** matrix, except for its symmetric and positive definite character. However, given suitable forms to the **W** matrix one can derive many updated Hessian formulae as the Powell,[12] MSP,[11,13,14] and others.[10,11]

Equation (10) indicates that the best general correction to the **B** matrix from a variational point of view is a rank two matrix. Applied to a quadratic function, because the vector **j** is conjugate to the previous directions, the **B**′ matrix will preserve the conjugancy condition if the **u** vector is also conjugate to the previous directions.[1,16] Equation (7) is a measure of the difference between **B**′ and **G**′. The value of this trace at the stationary point, $N_{\mathbf{W}}(\mathbf{E}^*)$, is[11]

$$N_{\mathbf{W}}(\mathbf{E}^*) = \frac{\mathbf{j}^T\mathbf{W}\mathbf{j}}{\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x}}\left[2 - \frac{(\Delta \mathbf{x}^T\mathbf{j})^2}{\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x}\mathbf{j}^T\mathbf{W}\mathbf{j}}\right]. \tag{11}$$

In many situations the **W** matrix and, consequently, the **M** matrix depend on a parameter,[11] say ϕ. The optimum value of this parameter that minimizes $N_{\mathbf{W}}(\mathbf{E})$ can be obtained by derivation of eq. (9) at the stationary point $\mathbf{E}^*, \mathbf{p}^*, \mathbf{L}^*$ with respect to ϕ and imposing the stationary condition; that is,

$$\frac{dN_{\mathbf{W}}(\mathbf{E}^*)}{d\phi} = \frac{d\Phi(\mathbf{E}(\phi)^*, \mathbf{p}(\phi)^*, \mathbf{L}(\phi)^*, \mathbf{W}(\phi))}{d\phi} = \frac{\partial \mathbf{E}^T}{\partial \phi}\nabla_{\mathbf{E}}\Phi + \frac{\partial \mathbf{p}^T}{\partial \phi}\nabla_{\mathbf{p}}\Phi + \frac{\partial \mathbf{L}^T}{\partial \phi}\nabla_{\mathbf{L}}\Phi + \frac{\partial \Phi}{\partial \phi} = 0. \tag{12}$$

Using the fact that Φ is stationary with respect to **E**, **p**, and **L** at the point $\mathbf{E}^*$, $\mathbf{p}^*$, and $\mathbf{L}^*$, one gets after some rearrangements

$$\mathbf{j}^T\mathbf{W}\mathbf{M}'\mathbf{W}\mathbf{j}(\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x})^2 + \left[\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x}\mathbf{j}^T\mathbf{W}\mathbf{j} - (\Delta \mathbf{x}^T\mathbf{j})^2\right]\Delta \mathbf{x}^T\mathbf{M}'\Delta \mathbf{x} = 0, \tag{13}$$

where **M**′ is the derivative of the **M** matrix with respect to the ϕ parameter. In the deduction of expression (13) the derivative of $\mathbf{WM} = \mathbf{I}$ was used. Note that the term in brackets is positive or equal to zero due to the Cauchy–Schwarz inequality. Given a $\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x}$ and $\mathbf{j}^T\mathbf{W}\mathbf{j}$ different from zero, if $[\Delta \mathbf{x}^T\mathbf{M}\Delta \mathbf{x}\mathbf{j}^T\mathbf{W}\mathbf{j} - (\Delta \mathbf{x}^T\mathbf{j})^2] \neq 0$ and the **M**′ matrix is positive definite, the stationary condition is satisfied by the condition $\mathbf{j}^T\mathbf{W}\mathbf{M}'\mathbf{W}\mathbf{j} = \Delta \mathbf{x}^T\mathbf{M}'\Delta \mathbf{x} = 0$. On the other hand, if the term in brackets is zero, then the stationary condition should be satisfied by the single condition $\mathbf{j}^T\mathbf{W}\mathbf{M}'\mathbf{W}\mathbf{j} = 0$. However, as will be seen below, these results have limited practical value.

DIRECT VARIATIONAL DEDUCTION OF BFGS FORMULA TO UPDATE B MATRIX

From the author's knowledge, all variational deductions of the BFGS formula to update the **B** matrix were obtained indirectly. First the correction to the **H** matrix is obtained using a formula like (10) for inverse updating and after that one

evaluates its inverse.[1,24,25] Here we present a direct variational deduction based on expression (10). Because the **E** matrix represents a variation of the **B** matrix, the best **W** matrix should be an approximation to the next average integral,[24]

$$\mathbf{W} = \frac{\int_0^1 \mathbf{G}^{-1}(\mathbf{x}^k + a\Delta\mathbf{x})\, da}{\int_0^1 da}. \tag{14}$$

A reasonable approximation to this average integral consists of using the **H** and **H**′ matrices rather than the full variation of the true inverse Hessian matrix, $\mathbf{G}^{-1}$, and replacing the integral by a discrete summation

$$\mathbf{W} = \frac{a\mathbf{H}' + b\mathbf{H}}{a+b} \quad a \geq 0, \quad b \geq 0, \quad a + b = 1. \tag{15}$$

In this way the **M** matrix should be an approximation of the average integral of the true Hessian matrix **G**; that is,

$$\mathbf{M} = a\mathbf{B}' + b\mathbf{B} \quad a \geq 0, \quad b \geq 0, \quad a + b = 1. \tag{16}$$

Note that the **M** matrix is positive definite if **B** and **B**′ are positive definite. In order to evaluate the parameters a and b, we impose the following condition:

$$\begin{aligned} \Delta\mathbf{x}^T\mathbf{M}\Delta\mathbf{x} &= a\Delta\mathbf{x}^T\Delta\mathbf{g} + b\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x} \\ &= (\Delta\mathbf{x}^T\Delta\mathbf{g}\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x})^{1/2} \\ a \geq 0, \quad b \geq 0, \quad a + b &= 1, \end{aligned} \tag{17}$$

where eq. (6) has been used. Now from eq. (17) and the fact that $a + b = 1$, we get

$$\begin{aligned} a &= \frac{\left(\frac{\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}}{\Delta\mathbf{x}^T\Delta\mathbf{g}}\right)^{1/2}}{1 + \left(\frac{\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}}{\Delta\mathbf{x}^T\Delta\mathbf{g}}\right)^{1/2}} \\ &= \frac{(\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x})^{1/2}}{(\Delta\mathbf{x}^T\Delta\mathbf{g})^{1/2} + (\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x})^{1/2}}, \end{aligned} \tag{18a}$$

$$\begin{aligned} b &= \frac{1}{1 + \left(\frac{\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}}{\Delta\mathbf{x}^T\Delta\mathbf{g}}\right)^{1/2}} \\ &= \frac{(\Delta\mathbf{x}^T\Delta\mathbf{g})^{1/2}}{(\Delta\mathbf{x}^T\Delta\mathbf{g})^{1/2} + (\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x})^{1/2}}. \end{aligned} \tag{18b}$$

Substituting eqs. (18) into eq. (16) and the result into eq. (10), we obtain eq. (3) after some algebraic manipulations. From this deduction we can very easily analyze the behavior of the BFGS formula. Because the **M** matrix should be positive definite, then from eq. (16) it is necessary that the **B** and **B**′ matrices both be positive definite too. This means that the optimum situation for BFGS updating of the Hessian matrix is when the QNR algorithm is applied to minimize a function. In other conditions, such as in the optimization of a TS, the BFGS update formula presents a degradation because in this case the **M** matrix is not positive definite. This means that from the present variational theory one cannot use eq. (3) directly to optimize a TS. However, other important conclusions can be drawn from this deduction. First, BFGS is based on an **M** or **W** matrix that is very close to the correct average of the variation on the $\mathbf{G}^{-1}$ matrix. Second, the parameters a and b are functions of the quotient of two Rayleigh quotients, namely $\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x}/\Delta\mathbf{x}^T\Delta\mathbf{x}$ and $\Delta\mathbf{x}^T\mathbf{B}'\Delta\mathbf{x}/\Delta\mathbf{x}^T\Delta\mathbf{x} = \Delta\mathbf{x}^T\Delta\mathbf{g}/\Delta\mathbf{x}^T\Delta\mathbf{x}$, which take into account the behavior of **B** and **B**′. These two points explain the good performance generally observed with the BFGS update formula when it is applied to the optimization of a minimum. Finally, if in eq. (16) one forces $a = 1$, then $\mathbf{M} = \mathbf{B}'$, and substituting it in eq. (10) we get the DFP formula.[10,11] On the other hand, forcing $b = 1$, then $\mathbf{M} = \mathbf{B}$, gives the dual Greenstadt formula.[10,11] Using these results we conclude that the BFGS formula is an average of the DFP and dual Greenstadt update formulae.

TS-BFGS UPDATE HESSIAN FORMULA

From the above analysis of the BFGS formula, we can consider the possibility of adequately modifying this formula to apply it to saddle point optimizations. The first serious and maybe the single problem is that the **M** matrix is nonpositive defined, as pointed out before. In BFGS the **u** vector is

$$\mathbf{u} = \frac{1}{(\Delta\mathbf{g}^T\Delta\mathbf{x}\Delta\mathbf{x}^T\mathbf{B}\Delta\mathbf{x})^{1/2}}(a\Delta\mathbf{g} + b\mathbf{B}\Delta\mathbf{x}), \tag{19}$$

which is clearly a vector that is a linear combination of the $\Delta\mathbf{g}$ and $\mathbf{B}\Delta\mathbf{x}$ vectors. Now, if we define the **M** matrix in the following way,

$$\mathbf{M} = (1 - \phi)|\mathbf{B}| + \phi\Delta\mathbf{g}\Delta\mathbf{g}^T, \tag{20}$$

where the $|\mathbf{B}|$ is the $\mathbf{B}$ matrix, but forcing all eigenvalues to be positive, that is,

$$|\mathbf{B}| = \sum_{i=1}^{n} |\lambda_i| \mathbf{v}_i \mathbf{v}_i^T, \tag{21}$$

where $\{\lambda_i, \mathbf{v}_i\}_{i=1}^{n}$ are the eigenpairs of the $\mathbf{B}$ matrix. Equation (20) can be seen as a rank one correction to the $(1 - \phi)|\mathbf{B}|$ matrix. Note that the $\mathbf{M}$ matrix defined in this way is positive definite if $1 \geq \phi \geq 0$. The corresponding $\mathbf{W}$ matrix is

$$\mathbf{W} = \frac{1}{(1-\phi)} \times \left[|\mathbf{H}| - \frac{\phi}{1 - (1 - \Delta\mathbf{g}^T|\mathbf{H}|\Delta\mathbf{g})\phi} |\mathbf{H}|\Delta\mathbf{g}\Delta\mathbf{g}^T|\mathbf{H}| \right], \tag{22}$$

where $|\mathbf{H}| = (|\mathbf{B}|)^{-1}$. Now the $\mathbf{u}$ vector takes the following form

$$\mathbf{u} = \frac{1}{\left\{ \Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x} + \left[(\Delta\mathbf{x}^T\Delta\mathbf{g})^2 - \Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x} \right] \phi \right\}} \times [(1-\phi)|\mathbf{B}|\Delta\mathbf{x} + \phi(\Delta\mathbf{x}^T\Delta\mathbf{g})\,\Delta\mathbf{g}], \tag{23}$$

which is a linear combination of $\mathbf{B}\Delta\mathbf{x}$ with $\mathbf{B}$ forced to be positive definite and the $\Delta\mathbf{g}$ vectors as in the BFGS formula. Substituting eq. (23) into eq. (10) we get a new update formula, labeled $\mathbf{B}'_{\text{TS-BFGS}}$; that is,

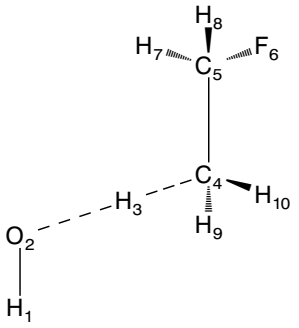
$$\begin{aligned} \mathbf{B}'_{\text{TS-BFGS}} &= \mathbf{B} + \frac{(\Delta\mathbf{g} - \mathbf{B}\Delta\mathbf{x})[(1-\phi)|\mathbf{B}|\Delta\mathbf{x} + \phi(\Delta\mathbf{x}^T\Delta\mathbf{g})\,\Delta\mathbf{g}]^T}{\{\Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x} + [(\Delta\mathbf{x}^T\Delta\mathbf{g})^2 - \Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x}]\phi\}} \\ &+ \frac{[(1-\phi)|\mathbf{B}|\Delta\mathbf{x} + \phi(\Delta\mathbf{x}^T\Delta\mathbf{g})\,\Delta\mathbf{g}](\Delta\mathbf{g} - \mathbf{B}\Delta\mathbf{x})^T}{\{\Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x} + [(\Delta\mathbf{x}^T\Delta\mathbf{g})^2 - \Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x}]\phi\}} \\ &- \frac{\Delta\mathbf{x}^T(\Delta\mathbf{g} - \mathbf{B}\Delta\mathbf{x})}{\{\Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x} + [(\Delta\mathbf{x}^T\Delta\mathbf{g})^2 - \Delta\mathbf{x}^T|\mathbf{B}|\Delta\mathbf{x}]\phi\}^2} \\ &\times [(1-\phi)|\mathbf{B}|\Delta\mathbf{x} + \phi(\Delta\mathbf{x}^T\Delta\mathbf{g})\,\Delta\mathbf{g}] \\ &\times [(1-\phi)|\mathbf{B}|\Delta\mathbf{x} + \phi(\Delta\mathbf{x}^T\Delta\mathbf{g})\,\Delta\mathbf{g}]^T. \end{aligned} \tag{24}$$

To obtain the best ϕ parameter one should substitute eqs. (20) and (22) and the derivative of the $\mathbf{M}$ matrix with respect to ϕ into eq. (13). However, this gives a very complicated equation on ϕ to be solved and, in addition, sometimes the value of the resulting ϕ does not fall in the domain $1 \geq \phi \geq 0$. Instead of this we selected the parameter ϕ as the square cosine of the angle between the $\Delta\mathbf{x}$ and $\mathbf{j}$ vectors. In this way the ϕ parameter falls in the correct domain.

PRACTICAL IMPLEMENTATION OF TS-BFGS UPDATE HESSIAN FORMULA

The update Hessian formula just derived is adequate for any algorithm of the type QA/TRIM or RQNR briefly described above with the eigenvector following method.[20] This is because one needs

TABLE I.
Starting and Final Geometrical Parameters for First Transition Structure of Reaction $CH_3CH_2F + OH^- \rightarrow CH_2{=}CH_2 + H_2O + F^-$.



Parameter	Starting	Final
H_1O_2	0.944	0.943
O_2H_3	1.500	1.559
H_3C_4	1.254	1.222
C_4C_5	1.472	1.503
C_5F_6	1.600	1.393
C_5H_7	1.112	1.126
C_5H_8	1.112	1.126
C_4H_9	1.110	1.111
C_4H_{10}	1.110	1.111
$H_3O_2H_1$	101.8	101.8
$C_4H_3O_2$	177.3	178.8
$C_5C_4H_3$	108.2	108.4
$F_6C_5C_4$	111.2	114.0
$H_7C_5C_4$	114.0	109.9
$H_8C_5C_4$	114.0	109.9
$H_9C_4C_5$	112.2	111.2
$H_{10}C_4C_5$	112.2	111.2
$H_1O_2H_3C_4$	0.0	5.0
$O_2H_3C_4C_5$	4.8	162.0
$H_3C_4C_5F_6$	180.0	180.0
$H_7C_5C_4F_6$	115.0	120.6
$H_8C_5C_4F_6$	−115.0	−120.6
$H_9C_4C_5H_3$	118.2	119.3
$H_{10}C_4C_5H_3$	−118.2	−119.2

Bond lengths in angstroms and bond angles and dihedrals in degrees.

the full diagonalization of the current **B** matrix. Rather than evaluating $\mathbf{B}'_{\text{TS-BFGS}}$ by using eq. (24), we employ the following algorithm:

Reset the vector $\mathbf{u} = \mathbf{0}$.

Compute $\Delta\mathbf{g}$ and store it in the **u** vector. Compute the scalar product $\mathbf{u}^T\Delta\mathbf{x}$.

Do the matrix–vector product $\mathbf{B}\Delta\mathbf{x}$ and store it in the **j** vector. Build $\mathbf{j} \leftarrow \mathbf{u} - \mathbf{j}$.

Compute the scalar products $\mathbf{j}^T\Delta\mathbf{x}$, $\Delta\mathbf{x}^T\Delta\mathbf{x}$, and $\mathbf{j}^T\mathbf{j}$. Evaluate $\phi = (\mathbf{j}^T\Delta\mathbf{x})^2/(\Delta\mathbf{x}^T\Delta\mathbf{x}\mathbf{j}^T\mathbf{j})$.

Compute $\mathbf{u} \leftarrow \phi\mathbf{u}(\mathbf{u}^T\Delta\mathbf{x})$ and after that

$$\mathbf{u} \leftarrow \mathbf{u} + (1 - \phi)\sum_{i=1}^{n} |\lambda_i|\mathbf{v}_i(\mathbf{v}_i^T\Delta\mathbf{x}). \quad (25)$$

Compute the scalar product $\mathbf{u}^T\Delta\mathbf{x}$ and evaluate $\mathbf{u} \leftarrow \mathbf{u}(\mathbf{u}^T\Delta\mathbf{x})^{-1}$.

Finally, with the vectors **j** and **u** and the scalar product $\mathbf{j}^T\Delta\mathbf{x}$, compute the $\mathbf{B}'_{\text{TS-BFGS}}$ matrix using eq. (10).

TABLE II.
Behaviors of TS-BFGS, MSP, and Powell Update Formulae along Optimization Process of First Transition Structure for Reaction $CH_3CH_2F + OH^- \rightarrow CH_2{=}CH_2 + H_2O + F^-$.

	TS-BFGS		MSP			Powell
Iteration	$\|\mathbf{g}\|$[a]	ϕ[b]	$\|\mathbf{g}\|$[a]	$\cos^2\omega$[c]	α[d]	$\|\mathbf{g}\|$[a]
1	29.76	0.86	29.76	0.965	0.00[e]	29.76
2	27.34	0.90	27.34	0.963	0.00[e]	27.34
3	22.64	0.90	22.64	0.936	0.00[e]	22.64
4	13.49	0.59	13.48	0.599	0.00[e]	13.48
5	3.13	0.66	3.17	0.570	0.00[e]	3.12
6	0.71	0.02	0.55	0.073	0.00[e]	2.06
7	0.22	0.07	0.30	0.003	1.00	0.64
8	0.12	0.65	0.14	0.007	0.99	0.40
9	0.05	0.04	0.09	0.314	0.00[e]	0.34
10	0.01	0.00	0.03	0.401	0.00[e]	0.31
11	$5 \cdot 10^{-3}$	0.01	0.02	0.037	0.95	0.26
12	$5 \cdot 10^{-3}$	0.03	$4 \cdot 10^{-3}$	0.021	0.98	0.19
13	$6 \cdot 10^{-3}$	0.13	$3 \cdot 10^{-3}$	0.000	1.00	0.11
14	$7 \cdot 10^{-3}$	0.24	$3 \cdot 10^{-3}$	0.000	1.00	0.02[f]
15	$8 \cdot 10^{-3}$	0.45	$4 \cdot 10^{-3}$	0.000	1.00	
16	0.01	0.45	$4 \cdot 10^{-3}$	0.000	1.00	
17	0.01	0.54	$5 \cdot 10^{-3}$	0.005	0.99	
18	0.02	0.27	$6 \cdot 10^{-3}$	0.008	0.99	
19	0.02	0.04	$7 \cdot 10^{-3}$	0.019	0.99	
20	0.02	0.11	$8 \cdot 10^{-3}$	0.038	0.97	
21	0.03	0.10	0.01	0.046	0.95	
22	0.04	0.01	0.01	0.038	0.96	
23	0.02	0.00	0.02	0.046	0.96	
24	0.02	0.00	0.02	0.051	0.00[e]	
25	0.01	0.00	0.02	0.080	0.00[e]	
26	0.01	0.00	0.03	0.017	0.98	
27	$2 \cdot 10^{-3}$	0.00	0.04	0.011	0.99	
28	$6 \cdot 10^{-4}$	—	0.02	0.001	1.00	
29			0.02	0.000	1.00	
30			0.01	0.002	1.00	
31			0.02	—	—	

[a] The RMS gradient, $\|\mathbf{g}\| = (\mathbf{g}^T\mathbf{g}/n)^{1/2}$ (kcal/mol).
[b] The parameter ϕ is defined as $\phi = (\Delta\mathbf{x}^T\Delta\mathbf{j})^2/(\Delta\mathbf{x}^T\Delta\mathbf{x}\Delta\mathbf{j}^T\Delta\mathbf{j})$. See text for more details.
[c] The $\cos^2\omega$ is defined in eq. (27), where ω is the angle between the vectors **j** and $\Delta\mathbf{x}$. See text for more details.
[d] The parameter α is computed according to the equation $\alpha = \phi^2 = (1 - \cos^2\omega)$.
[e] In this iteration the strictly Murtagh–Sargent[22] update formula is used. See text for more details.
[f] Stationary point with two negative eigenvalues.

Discussion and Examples

First we comment on the correct implementation of the MSP update formula described in ref. 13 by considering the results that follow eq. (12). Briefly, the MSP update is a hybrid combination of the MS update[22] and the Powell update[12]; that is, $\mathbf{B}'_{\mathrm{MSP}(\alpha)} = (1-\alpha)\mathbf{B}'_{\mathrm{MS}} + \alpha\mathbf{B}'_{\mathrm{P}}$, where $\mathbf{B}'_{\mathrm{MP}}$ and $\mathbf{B}'_{\mathrm{P}}$ are the MS and Powell updates, respectively, and α is an arbitrary parameter defined in the domain $1 \geq \alpha \geq 0$. The $\mathbf{M}$ matrix that charactreizes the MSP update formula, $\mathbf{B}'_{\mathrm{MSP}(\alpha)}$, is[11]

$$\mathbf{M} = \phi\mathbf{I} + (1-\phi)\frac{\Delta\mathbf{x}^T\Delta\mathbf{x}}{(\Delta\mathbf{x}^T\mathbf{j})}\mathbf{j}\mathbf{j}^T \quad 1 \geq \phi \geq 0, \quad (26)$$

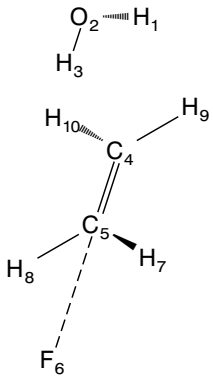
where $\phi^2 = \alpha$. Now the problem is to see which is the best ϕ parameter, that is, the ϕ parameter that minimizes $N_{\mathbf{W}}(\mathbf{E}^*)$. Evaluating the inverse matrix $\mathbf{W}$ and the derivative with respect to ϕ, $\mathbf{M}'$, corresponding to the $\mathbf{M}$ matrix defined in eq. (26), and substituting the results into eq. (13) we get

$$\phi\frac{(\Delta\mathbf{x}^T\mathbf{j})^2}{\Delta\mathbf{x}^T\Delta\mathbf{x}\,\mathbf{j}^T\mathbf{j}} = \phi\cos^2\omega = 0 \quad 1 \geq \phi \geq 0, \quad (27)$$

where ω is the angle formed between the vectors $\mathbf{j}$ and $\Delta\mathbf{x}$. Equation (27) tell us that when $\cos^2\omega \neq 0$, the optimum ϕ is $\phi = 0$, which corresponds to the MS update.[11] On the other hand, if $\cos^2\omega = 0$, any ϕ between 1 and 0 is, in principle, optimum. However, the last situation very often occurs near the convergence, but in this case the MS update is not the optimum.[1] Consequently, one should take any ϕ such that $1 \geq \phi > 0$. From a practical point of view, when $\cos^2\omega \leq \varepsilon$, ε being a small number, the parameter ϕ is selected as $\phi = (1-\cos^2\omega)^{1/2}$.[11,13] A suitable choice for ε is 0.05. Obviously, if along the optimization process $\cos^2\omega \leq \varepsilon$, the MSP update coincides with the Powell update because in this case $\phi \approx 1$, and, consequently, $\alpha \approx 1$.

Now we present the behavior and the performance of the BFGS, TS-BFGS, MSP, and Powell update formulae in the location and optimization of TSs of several reactions. The calculations were performed at the AM1[26] and *ab initio* Hartree–Fock levels. The optimizations were carried out in internal coordinates. The algorithms used to optimize the TSs were the RQNR described in ref. 13 for the AM1 calculations and the QA/TRIM algorithm[18] as implemented in the GAMESS package program[27] for the *ab initio* calculations. The AM1 calculations were using a modified version of the MOPAC program.[28] In the AM1 optimizations, the convergence criteria were taken from: the maximum component of $\Delta\mathbf{x}$ and its root mean square (RMS), $\|\Delta\mathbf{x}\| = (\Delta\mathbf{x}^T\Delta\mathbf{x}/n)^{1/2}$; the maximum component of $\mathbf{g}$ and its RMS, $\|\mathbf{g}\| = (\mathbf{g}^T\mathbf{g}/n)^{1/2}$; and

TABLE III.
Starting and Final Geometrical Parameters of Second Transition Structure for Reaction $CH_3CH_2F + OH^- \rightarrow CH_2{=}CH_2 + H_2O + F^-$.



Parameter	Starting	Final
H_1O_2	0.962	0.963
O_2H_3	0.966	0.965
H_3C_4	2.433	2.534
C_4C_5	1.334	1.336
C_5F_6	2.500	2.420
C_5H_7	1.100	1.100
C_5H_8	1.100	1.104
C_4H_9	1.097	1.096
C_4H_{10}	1.096	1.096
$H_3O_2H_1$	102.5	102.4
$C_4H_3O_2$	111.5	106.2
$C_5C_4H_3$	81.3	74.5
$F_6C_5C_4$	137.7	129.9
$H_7C_5C_4$	123.5	122.4
$H_8C_5C_4$	123.7	124.0
$H_9C_4C_5$	122.8	123.0
$H_{10}C_4C_5$	122.7	122.4
$H_1O_2H_3C_4$	0.0	−5.0
$O_2H_3C_4C_5$	179.3	156.5
$H_3C_4C_5F_6$	178.7	144.9
$H_7C_5C_4F_6$	87.6	99.9
$H_8C_5C_4F_6$	−85.8	−74.2
$H_9C_4C_5H_3$	92.6	116.2
$H_{10}C_4C_5H_3$	−91.4	−68.5

Bond lengths in angstroms and bond angles and dihedrals in degrees.

the absolute value of the energy differences between two consecutive iterations $|f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)|$; with the threshold values $1.8 \cdot 10^{-3}$ Å, $1.2 \cdot 10^{-3}$ Å, $1.5 \cdot 10^{-1}$ kcal/mol Å, $7.5 \cdot 10^{-2}$ kcal/mol Å, and $5.0 \cdot 10^{-4}$ kcal/mol, respectively, which are the units used in the MOPAC program. In the case of the *ab initio* optimizations, the standard convergence criteria of the GAMESS program were used. In the GAMESS program a value of $7 \cdot 10^{-3}$ for the ε parameter was taken.

ELIMINATION REACTION $CH_3CH_2F + OH^- \rightarrow CH_2 = CH_2 + H_2O + F^-$

This reaction was studied by Cummins and Gready[29] with the AM1 method.[26] The wave function employed was the restricted Hartree–Fock. The overall reaction occurs through two transition structures as found for other elimination reactions.[30] In Table I we present the geometrical parameters of the initial and optimized molecular geometries calculated for the first TS. In Table II we compare the behavior of the optimization process using different update formulae. We observe that using the TS-BFGS update Hessian formula the algorithm converges within 28 iterations. Using the Powell update formula, the algorithm converges within 14 iterations. However, the stationary point reached possesses two imaginary frequencies, $331.3i$ cm^{-1} and $39.5i$ cm^{-1}. On the other hand, if we use the MSP update formula the algorithm needs 31 iterations to converge to the correct TS. The correct optimized geometrical parameters are those shown in Table I. At the final stationary point, the imaginary frequency $305.3i$ cm^{-1} is associated with the true transition vector. Finally we note that using the standard BFGS formula, eq. (3), the algorithm diverges.

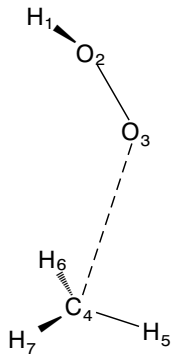
TABLE IV.
Behavior of TS-BFGS, MSP, and Powell Update Formulae along Optimization Process of Second Transition Structure for Reaction $CH_3CH_2F + OH^- \rightarrow CH_2{=}CH_2 + H_2O + F^-$.

	TS-BFGS		MSP			Powell
Iteration	$\|\mathbf{g}\|$[a]	ϕ[b]	$\|\mathbf{g}\|$[a]	$\cos^2\omega$[c]	α[d]	$\|\mathbf{g}\|$[a]
1	0.34	0.25	0.34	0.043	0.96	0.34
2	0.10	0.11	0.10	0.017	0.98	0.10
3	0.05	0.36	0.05	0.014	0.99	0.05
4	0.08	0.42	0.08	0.015	0.99	0.08
5	0.12	0.00	0.12	0.192	0.00[e]	0.12
6	0.15	0.52	0.16	0.512	0.00[e]	0.16
7	0.17	0.20	0.19	0.001	1.00	0.20
8	0.14	0.01	0.09	0.121	0.00[e]	0.09
9	0.04	0.24	0.03	0.025	0.97	0.03
10	0.03	0.27	0.02	0.002	1.00	0.02
11	0.02	0.22	0.01	0.001	1.00	0.02
12	0.02	0.23	0.01	0.000	1.00	0.02
13	0.01	0.21	0.01	0.000	1.00	0.01
14	0.01	0.40	0.01	0.000	1.00	0.01
15	0.01	0.68	0.01	0.000	1.00	0.02
16	0.01	0.50	0.01	0.000	1.00	0.02
17	0.01	0.19	0.02	0.000	1.00	0.03
18	$4 \cdot 10^{-3}$	0.04	0.06	0.002	1.00	0.05
19	$4 \cdot 10^{-3}$	0.01	0.04	—	—	0.02
20	$5 \cdot 10^{-3}$	0.00				0.02
21	$6 \cdot 10^{-3}$	0.01				
22	0.01	0.00				
23	0.02	0.01				
24	0.04	0.44				
25	0.04	—				

[a] The RMS gradient, $\|\mathbf{g}\| = (\mathbf{g}^T\mathbf{g}/n)^{1/2}$ (kcal/mol).
[b] The parameter ϕ is defined as $\phi = (\Delta\mathbf{x}^T\Delta\mathbf{j})^2/(\Delta\mathbf{x}^T\Delta\mathbf{x}\Delta\mathbf{j}^T\Delta\mathbf{j})$. See text for more details.
[c] The $\cos^2\omega$ is defined in eq. (27), where ω is the angle between the vectors $\mathbf{j}$ and $\Delta\mathbf{x}$. See text for more details.
[d] The parameter α is computed according to the equation $\alpha = \phi^2 = (1 - \cos^2\omega)$.
[e] In this iteration the strictly Murtagh–Sargent[22] update formula is used. See text for more details.

The results of the second TS are presented in Tables III and IV. In Table III we show the geometrical parameters. The value of the imaginary frequency is $240.9i$ cm^{-1}. In Table IV we show the behavior of the algorithm using the different updates of the Hessian matrix. The TS-BFGS update needs 25 iterations to converge, while the Powell and MSP ones need 20 and 19 iterations, respectively. For the MSP update, at iterations 5, 6, and 8 the α parameter is zero, which means that in this case MSP is just the MS update, while in the rest of the optimization process MSP is just the Powell update. We emphasize that using the three different update formulae the optimization process converges to the same molecular geometry with insignificant differences. Finally, the process does not converge employing the standard BFGS update formula.

TABLE V.
Starting and Final Geometrical Parameters of Transition Structure for Reaction $CH_3OOH \rightarrow CH_3 + OOH$.



Parameter	Starting	Final
H_1O_2	1.010	1.009
O_2O_3	1.177	1.178
O_3C_4	2.800	2.683
C_4H_5	1.086	1.087
C_4H_6	1.086	1.086
C_4H_7	1.088	1.086
$H_1O_2O_3$	112.6	112.5
$O_2O_3C_4$	119.5	117.3
$O_3C_4H_5$	70.0	68.8
$O_3C_4H_6$	92.3	75.9
$O_3C_4H_7$	107.8	125.0
$H_1O_2O_3C_4$	88.2	90.4
$O_2O_3C_4H_5$	174.5	177.5
$H_5O_3C_4H_6$	−120.9	−127.5
$H_5O_3C_4H_7$	116.8	116.0

Bond lengths in angstroms and bond angles in degrees.

DECOMPOSITION REACTION $CH_3OOH \rightarrow CH_3 + OOH$

This reaction was studied using the unrestricted Hartree–Fock version of the semiempirical AM1 Hamiltonian. In Table V we show the geometrical parameters and in Table VI the behaviors of the different update Hessian formulae. Using the TS-BFGS update, the algorithm needs 23 iterations to converge and the ϕ parameter is close to zero in almost all iterations. Regarding the α parameter, we observe that the MSP update is a genuine MS update in eight iterations. In this example the Powell update is the more efficient. Again, using the three different update formulae the optimization process converges to the same molecular geometry with insignificant differences. Finally, using the standard BFGS equation the algorithm converges to a minimum corresponding to the products of the reaction.

DISSOCIATION REACTION OF $H_2CO \rightarrow H_2 + CO$

This reaction was studied by Cerjan and Miller[31] at the Hartee–Fock *ab initio* level with the STO-2G basis set. We report for this reaction results obtained at the same level of theory using different update Hessian formulae. In Table VII we present the geometrical parameters for the initial and final geometry. Table VIII shows the behavior of the different update formulae. The Hessian matrix at the initial geometry is positive definite and the potential transition vector corresponds to the second eigenvector if the eigenpairs are in increasing value order. Employing the TS-BFGS update formula, in the first four iterations the Hessian matrix does not have the desired eigenvalue spectra. For the MSP update, in the first four iterations and at iteration 11 the Hessian matrix is positive definite. Also, in the Powell update at the first four iterations, the Hessian matrix does not present the correct structure. Using the three different update formulae the optimization process converges to the same molecular geometry with insignificant differences. Finally, and employing the standard BFGS formula, the process does not converge.

METHOXY RADICAL ISOMERIZATION, $CH_3O \rightarrow CH_2OH$

This reaction was studied by Culot et al.[18b] at the Hartree–Fock *ab initio* level. Here we report the results for the same reaction within the C_s

TABLE VI.
Behavior of TS-BFGS, MSP, and Powell Update Formulae along Optimization Process of Transition Structure for Reaction $CH_3OOH \rightarrow CH_3 + OOH$.

	TS-BFGS		MSP			Powell
Iteration	$\|\mathbf{g}\|$[a]	ϕ[b]	$\|\mathbf{g}\|$[a]	$\cos^2\omega$[c]	α[d]	$\|\mathbf{g}\|$[a]
1	0.62	0.14	0.62	0.136	0.00[e]	0.62
2	0.20	0.17	0.20	0.197	0.00[e]	0.20
3	0.14	0.18	0.14	0.239	0.00[e]	0.14
4	0.11	0.03	0.09	0.013	0.97	0.08
5	0.15	0.00	0.15	0.004	0.99	0.13
6	0.05	0.00	0.21	0.009	0.98	0.14
7	0.06	0.00	0.38	0.522	0.00[e]	0.19
8	0.04	0.00	0.16	0.167	0.00[e]	0.43
9	0.08	0.00	0.12	0.090	0.00[e]	0.20
10	0.07	0.01	0.09	0.009	0.98	0.08
11	0.17	0.60	0.07	0.006	0.99	0.07
12	0.11	0.07	0.07	0.019	0.96	0.06
13	0.10	0.07	0.07	0.051	0.00[e]	0.06
14	0.09	0.21	0.06	0.000	1.00	0.04
15	0.09	0.46	0.06	0.057	0.00[e]	0.05
16	0.08	0.03	0.06	0.030	0.94	0.04
17	0.07	0.00	0.06	0.008	0.98	0.04
18	0.07	0.00	0.05	0.000	1.00	0.02
19	0.07	0.01	0.04	0.000	1.00	
20	0.07	0.02	0.04	0.000	1.00	
21	0.07	0.11	0.03	0.000	1.00	
22	0.07	0.03	0.02	0.001	1.00	
23	0.07	—	0.02	0.002	1.00	
24			0.05	—	—	

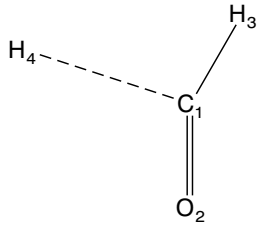
[a] The RMS gradient, $\|\mathbf{g}\| = (\mathbf{g}^T\mathbf{g}/n)^{1/2}$ (kcal/mol).
[b] The parameter ϕ is defined as $\phi = (\Delta\mathbf{x}^T\Delta\mathbf{j})^2/(\Delta\mathbf{x}^T\Delta\mathbf{x}\Delta\mathbf{j}^T\Delta\mathbf{j})$. See text for more details.
[c] The $\cos^2\omega$ is defined in eq. (27), where ω is the angle between the vectors $\mathbf{j}$ and $\Delta\mathbf{x}$. See text for more details.
[d] The parameter α is computed according to the equation $\alpha = \phi^2 = (1 - \cos^2\omega)$.
[e] In this iteration the strictly Murtagh–Sargent[22] update formula is used. See text for more details.

symmetry using the unrestricted Hartree–Fock wave function with the STO-3G basis set. In Table IX we show the geometry and in Table X the behavior of the three updated Hessians. The initial Hessian matrix is positive definite, but just after the first correction the Hessian achieves the correct eigenvalue spectra in all three cases. The three update formulae converge to the same molecular geometry. Using the standard BFGS formula, eq. (3), the algorithm diverges.

Concluding Remarks

We presented a BFGS-like updated Hessian formula to locate TSs. This TS-BFGS modification of the standard BFGS formula is understood in the sense that the $\mathbf{u}$ vector, which is present in any updating Hessian formula, is in this case a function of both the $\mathbf{B}\Delta\mathbf{x}$ and $\Delta\mathbf{g}$ vectors, with the $\mathbf{B}$ matrix forced to be positive definite as in the standard BFGS formula. This update presents some stability and efficiency and is quite competitive with respect to the normal update formulae used to locate TSs, such as the Powell and the MSP. Finally, we make the following consideration: because the BFGS update formula is the best formula to update Hessian matrices for a minimization algorithm, the TS-BFGS can be seen as its analogue for optimizing saddle points, and the TS-BFGS presents the same performance as the Powell or MSP update formulae, it is likely that the Powell and MSP formulae are the best update that can be formulated to optimize saddle points.

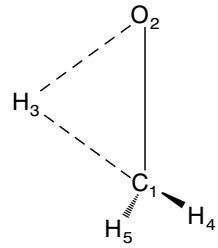
TABLE VII.
Starting and Final Geometrical Parameters of Transition Structure for Reaction $H_2CO \rightarrow H_2 + CO$.



Parameter	Starting	Final
C_1O_2	1.220	1.199
C_1H_3	1.110	1.122
C_1H_4	1.110	1.537
$O_2C_1H_3$	123.4	156.7
$O_2C_1H_4$	122.4	107.9
$O_2C_1H_3H_4$[a]	0.0	0.0

Bond lengths in angstroms and bond angles in degrees.
[a]Out of plane angle coordinate, defined from bond O_2C_1 to plane $C_1H_3H_4$.

TABLE IX.
Starting and Final Geometrical Parameters of Transition Structure for Reaction $CH_3O \rightarrow CH_2OH$ within C_S Symmetry.



Parameter	Starting	Final
C_1O_2	1.423	1.423
C_1H_3	1.484	1.326
C_1H_4	1.087	1.088
$O_2C_1H_3$	42.7	51.8
$O_2C_1H_4$	117.5	117.5
$H_3O_2C_1H_4$	−105.4	−106.4

Bond lengths in angstroms and bond angles and dihedrals in degrees.

TABLE VIII.
Behavior of TS-BFGS, MSP, and Powell Update Formulae along Optimization Process of Transition Structure for Reaction $H_2CO \rightarrow H_2 + CO$.

	TS-BFGS		MSP			Powell
Iteration	$\|\mathbf{g}\|$[a]	ϕ[b]	$\|\mathbf{g}\|$[a]	$\cos^2\omega$[c]	α[d]	$\|\mathbf{g}\|$[a]
1	$2.1 \cdot 10^{-2}$[e]	$4.2 \cdot 10^{-3}$	$2.1 \cdot 10^{-2}$[e]	$4.2 \cdot 10^{-3}$	$9.9 \cdot 10^{-1}$	$2.1 \cdot 10^{-2}$[e]
2	$4.0 \cdot 10^{-2}$[e]	$6.4 \cdot 10^{-3}$	$4.0 \cdot 10^{-2}$[e]	$6.0 \cdot 10^{-3}$	$9.9 \cdot 10^{-1}$	$4.0 \cdot 10^{-2}$[e]
3	$6.5 \cdot 10^{-2}$[e]	$9.2 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$[e]	$8.6 \cdot 10^{-2}$	0.0[f]	$6.5 \cdot 10^{-2}$[e]
4	$8.5 \cdot 10^{-2}$[e]	$4.5 \cdot 10^{-1}$	$8.4 \cdot 10^{-2}$[e]	$6.6 \cdot 10^{-1}$	0.0[f]	$8.5 \cdot 10^{-2}$[e]
5	$9.5 \cdot 10^{-2}$	$8.4 \cdot 10^{-1}$	$9.8 \cdot 10^{-2}$	$7.3 \cdot 10^{-1}$	0.0[f]	$9.5 \cdot 10^{-2}$
6	$1.3 \cdot 10^{-1}$	$6.0 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$	$6.4 \cdot 10^{-1}$	0.0[f]	$1.2 \cdot 10^{-1}$
7	$1.0 \cdot 10^{-1}$	$7.3 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$7.2 \cdot 10^{-1}$	0.0[f]	$1.0 \cdot 10^{-1}$
8	$8.6 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	$8.5 \cdot 10^{-2}$	$3.0 \cdot 10^{-1}$	0.0[f]	$8.8 \cdot 10^{-2}$
9	$4.4 \cdot 10^{-2}$	$2.1 \cdot 10^{-1}$	$4.2 \cdot 10^{-2}$	$4.3 \cdot 10^{-1}$	0.0[f]	$4.7 \cdot 10^{-2}$
10	$5.2 \cdot 10^{-2}$	$4.2 \cdot 10^{-1}$	$3.7 \cdot 10^{-2}$	$4.3 \cdot 10^{-1}$	0.0[f]	$5.1 \cdot 10^{-2}$
11	$1.9 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$[e]	$6.3 \cdot 10^{-2}$	0.0[f]	$3.1 \cdot 10^{-2}$
12	$1.0 \cdot 10^{-2}$	$8.4 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$5.5 \cdot 10^{-1}$	0.0[f]	$1.2 \cdot 10^{-2}$
13	$6.1 \cdot 10^{-3}$	$9.6 \cdot 10^{-2}$	$2.1 \cdot 10^{-3}$	0.0	1.0	$6.5 \cdot 10^{-3}$
14	$2.2 \cdot 10^{-3}$	$4.3 \cdot 10^{-2}$	$9.9 \cdot 10^{-4}$	$1.4 \cdot 10^{-1}$	0.0[f]	$4.6 \cdot 10^{-3}$
15	$2.4 \cdot 10^{-4}$	$1.7 \cdot 10^{-1}$	$1.0 \cdot 10^{-4}$	$1.2 \cdot 10^{-3}$	1.0	$1.0 \cdot 10^{-3}$
16	$1.5 \cdot 10^{-4}$	$3.8 \cdot 10^{-1}$	$8.0 \cdot 10^{-6}$	—	—	$6.5 \cdot 10^{-4}$
17	$1.0 \cdot 10^{-5}$	—				$1.2 \cdot 10^{-4}$
18						$7.9 \cdot 10^{-6}$

[a]The RMS gradient, $\|\mathbf{g}\| = (\mathbf{g}^T\mathbf{g}/n)^{1/2}$ (Hartree/Å).
[b]The parameter ϕ is defined as $\phi = (\Delta\mathbf{x}^T\Delta\mathbf{j})^2/(\Delta\mathbf{x}^T\Delta\mathbf{x}\Delta\mathbf{j}^T\Delta\mathbf{j})$. See text for more details.
[c]The $\cos^2\omega$ is defined in eq. (27), where ω is the angle between the vectors $\mathbf{j}$ and $\Delta\mathbf{x}$. See text for more details.
[d]The parameter α is computed according to the equation $\alpha = \phi^2 = (1 - \cos^2\omega)$.
[e]Iterations where the Hessian matrix does not have the correct spectra.
[f]In this iteration the strictly Murtagh–Sargent[22] update formula is used. See text for more details.

TABLE X.
Behavior of TS-BFGS, MSP, and Powell Update Formulae along Optimization Process of Transition Structure for Reaction $CH_3O \rightarrow CH_2OH$.

	TS-BFGS		MSP			Powell
Iteration	$\|\mathbf{g}\|$[a]	ϕ[b]	$\|\mathbf{g}\|$[a]	$\cos^2\omega$[c]	α[d]	$\|\mathbf{g}\|$[a]
1	$3.6 \cdot 10^{-2}$	$3.0 \cdot 10^{-3}$	$3.6 \cdot 10^{-2}$	$3.0 \cdot 10^{-3}$	$9.9 \cdot 10^{-1}$	$3.6 \cdot 10^{-2}$
2	$4.2 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$3.6 \cdot 10^{-2}$	$1.9 \cdot 10^{-1}$	0.0[e]	$3.6 \cdot 10^{-2}$
3	$1.0 \cdot 10^{-1}$	$3.7 \cdot 10^{-3}$	$4.2 \cdot 10^{-2}$	$5.6 \cdot 10^{-1}$	0.0[e]	$4.9 \cdot 10^{-2}$
4	$7.8 \cdot 10^{-2}$	$6.6 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	0.0[e]	$3.7 \cdot 10^{-2}$
5	$3.0 \cdot 10^{-2}$	$4.1 \cdot 10^{-2}$	$1.9 \cdot 10^{-3}$	$1.7 \cdot 10^{-2}$	0.0[e]	$2.3 \cdot 10^{-2}$
6	$4.2 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$8.0 \cdot 10^{-1}$	0.0[e]	$4.1 \cdot 10^{-3}$
7	$8.9 \cdot 10^{-3}$	$3.9 \cdot 10^{-1}$	$6.6 \cdot 10^{-3}$	$6.1 \cdot 10^{-1}$	0.0[e]	$5.5 \cdot 10^{-4}$
8	$1.2 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$1.1 \cdot 10^{-3}$	$3.6 \cdot 10^{-2}$	0.0[e]	$1.4 \cdot 10^{-4}$
9	$1.0 \cdot 10^{-3}$	$9.0 \cdot 10^{-2}$	$4.3 \cdot 10^{-5}$	$5.3 \cdot 10^{-1}$	0.0[e]	$8.3 \cdot 10^{-5}$
10	$4.6 \cdot 10^{-4}$	$1.6 \cdot 10^{-2}$	$1.5 \cdot 10^{-5}$	—	—	$5.9 \cdot 10^{-6}$
11	$6.7 \cdot 10^{-5}$	$2.3 \cdot 10^{-1}$				
12	$9.1 \cdot 10^{-6}$	—				

[a]The RMS gradient, $\|\mathbf{g}\| = (\mathbf{g}^T\mathbf{g}/n)^{1/2}$ (Hartree/Å).
[b]The parameter ϕ is defined as $\phi = (\Delta\mathbf{x}^T\Delta\mathbf{j})^2/(\Delta\mathbf{x}^T\Delta\mathbf{x}\Delta\mathbf{j}^T\Delta\mathbf{j})$. See text for more details.
[c]The $\cos^2\omega$ is defined in eq. (27), where ω is the angle between the vectors $\mathbf{j}$ and $\Delta\mathbf{x}$. See text for more details.
[d]The parameter α is computed according to equation $\alpha = \phi^2 = (1 - \cos^2\omega)$.
[e]In this iteration the strictly Murtagh–Sargent[22] update formula is used. See text for more details.

Appendix A: Inverse of Matrix B′ Given by Eq. (10)

We are looking for the matrix $\mathbf{H}'$ such that $\mathbf{B}'\mathbf{H}' = \mathbf{I}$. We assume that $\mathbf{BH} = \mathbf{I}$. Using the Sherman–Morrison formula[1] on eq. (10) we get

$$\mathbf{H}' = \mathbf{H} - \mathbf{H}[\mathbf{j} \quad \mathbf{u}]\mathbf{C}^{-1}[\mathbf{j} \quad \mathbf{u}]^T\mathbf{H}. \qquad \text{(A.1)}$$

In eq. (A.1) the $\mathbf{C}$ matrix is

$$\begin{aligned}\mathbf{C} &= \begin{bmatrix} 0 & 1 \\ 1 & -(\mathbf{j}^T\Delta\mathbf{x}) \end{bmatrix}^{-1} + [\mathbf{j} \quad \mathbf{u}]^T\mathbf{H}[\mathbf{j} \quad \mathbf{u}] \\ &= \begin{bmatrix} 0 & 1 \\ 1 & -(\mathbf{j}^T\Delta\mathbf{x}) \end{bmatrix}^{-1} + \begin{bmatrix} -\mathbf{j}^T\mathbf{n} & -\mathbf{u}^T\mathbf{n} \\ -\mathbf{u}^T\mathbf{n} & \mathbf{u}^T\mathbf{Hu} \end{bmatrix} \\ &= \begin{bmatrix} -\Delta\mathbf{g}^T\mathbf{n} & \mathbf{u}^T\mathbf{H}\Delta\mathbf{g} \\ \Delta\mathbf{g}^T\mathbf{Hu} & \mathbf{u}^T\mathbf{Hu} \end{bmatrix}, \end{aligned} \qquad \text{(A.2)}$$

where the following relations and definitions were used: $\mathbf{Hj} = \mathbf{H}\Delta\mathbf{g} - \Delta\mathbf{x} = -\mathbf{n}$, $\mathbf{j}^T\mathbf{n} = \mathbf{n}^T\Delta\mathbf{g} + \mathbf{j}^T\Delta\mathbf{x}$, and $\mathbf{u}^T\mathbf{n} = 1 - \mathbf{u}^T\mathbf{H}\Delta\mathbf{g}$. If $\mathbf{C}^{-1} = \mathbf{D}$, where the symmetric $\mathbf{D}$ matrix is

$$\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \qquad \text{(A.3)}$$

then the value of its elements are

$$D_{11} = -\frac{\mathbf{u}^T\mathbf{Hu}}{(\Delta\mathbf{g}^T\mathbf{Hu})^2 + \mathbf{u}^T\mathbf{Hu}\Delta\mathbf{g}^T\mathbf{n}}, \qquad \text{(A.4a)}$$

$$D_{12} = D_{21} = \frac{\Delta\mathbf{g}^T\mathbf{Hu}}{(\Delta\mathbf{g}^T\mathbf{Hu})^2 + \mathbf{u}^T\mathbf{Hu}\Delta\mathbf{g}^T\mathbf{n}}, \qquad \text{(A.4b)}$$

$$D_{22} = \frac{\Delta\mathbf{g}^T\mathbf{n}}{(\Delta\mathbf{g}^T\mathbf{Hu})^2 + \mathbf{u}^T\mathbf{Hu}\Delta\mathbf{g}^T\mathbf{n}}. \qquad \text{(A.4c)}$$

Now substituting eqs. (A.4) into eq. (A.3) and the result into eq. (A.1), we get the expression for $\mathbf{H}'$ in a condensed form,

$$\begin{aligned}\mathbf{H}' &= \mathbf{H} - D_{11}\mathbf{nn}^T + D_{12}(\mathbf{nu}^T\mathbf{H} + \mathbf{Hun}^T) \\ &\quad - D_{22}\mathbf{Huu}^T\mathbf{H} \\ &= \mathbf{H} + [\mathbf{n} \quad \mathbf{Hu}]\begin{bmatrix} -D_{11} & D_{12} \\ D_{21} & -D_{22} \end{bmatrix}[\mathbf{n} \quad \mathbf{Hu}]^T \\ &= \mathbf{H} + \mathbf{Q}. \end{aligned} \qquad \text{(A.5)}$$

Equation (A.5) tells us that the general correction to the $\mathbf{H}$ matrix is again a rank two matrix.

Acknowledgments

We are indebted to Professor S. Olivella for his valuable suggestions. This research was supported by the Spanish DGICYT (Grant PB95-0278-C02-01).

References

1. R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.
2. W. C. Davidon, *Variable Metric Method for Minimization*, AEC Research and Development Report, Argonne National Laboratory Report, ANL-5990, 1959.
3. R. Fletcher and M. J. D. Powell, *Comput. J.*, **6**, 149 (1963).
4. C. G. Broyden, *J. Inst. Math. Comput.*, **6**, 222 (1970).
5. R. Fletcher, *Comput. J.*, **13**, 317 (1970).
6. D. Goldfarb, *Math. Comput.*, **24**, 23 (1970).
7. D. F. Shanno, *Math. Comput.*, **24**, 647 (1970).
8. H. Y. Huang, *J. Opt. Theor. Applic.*, **5**, 405 (1970).
9. M. R. Hestenes, *Conjugate Directions Methods in Optimization*, Springer–Verlag, New York, 1980.
10. J. Greenstadt, *Math. Comput.*, **24**, 1 (1970).
11. J. M. Bofill and M. Comajuan, *J. Comput. Chem.*, **16**, 1326 (1995).
12. M. J. D. Powell, *Math. Prog.*, **1**, 26 (1971).
13. J. M. Bofill, *J. Comput. Chem.*, **15**, 1 (1994).
14. J. M. Bofill, *Chem. Phys. Lett.*, **260**, 359 (1996).
15. J. Simons, P. Jørgensen, H. Taylor, and J. Ozment, *J. Phys. Chem.*, **87**, 2745 (1983).
16. L. E. Scales, *Introduction to Non-Linear Optimization*, Macmillan, Basingstoke, U.K., 1985.
17. J. M. Anglada and J. M. Bofill, *Int. J. Quantum Chem.*, **62**, 153 (1997).
18. (a) T. Helgaker, *Chem. Phys. Lett.*, **182**, 503 (1991); (b) P. Culot, G. Dive, V. H. Nguyen, and J. M. Ghuysen, *Theor. Chim. Acta*, **82**, 189 (1992).
19. F. Jensen, *J. Chem. Phys.*, **102**, 6706 (1995).
20. A. Banderjee, N. Adams, J. Simons, and R. Shepard, *J. Phys. Chem.*, **89**, 52 (1985).
21. H. B. Schlegel, In *Modern Electronic Structure Theory*, D. R. Yarkony, Ed., World Scientific Publishing, Singapore, 1995.
22. B. A. Murtagh and R. W. H. Sargent, *Comput. J.*, **13**, 185 (1972).
23. H. B. Schlegel, *Theor. Chim. Acta*, **83**, 15 (1992).
24. M. Al-Baali and R. Fletcher, *J. Oper. Res. Soc.*, **36**, 405 (1985).
25. J. E. Dennis and R. E. Schnabel, *SIAM Rev.*, **21**, 443 (1979).
26. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
27. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.*, **14**, 1347 (1993).
28. *MOPAC Program*, local version.
29. P. L. Cummins and J. E. Gready, *J. Comput. Chem.*, **10**, 939 (1989).
30. N. S. Isaacs, *Physical Organic Chemistry*, Longman Scientific, New York, 1987.
31. C. J. Cerjan and W. H. Miller, *J. Chem. Phys.*, **75**, 2800 (1981).